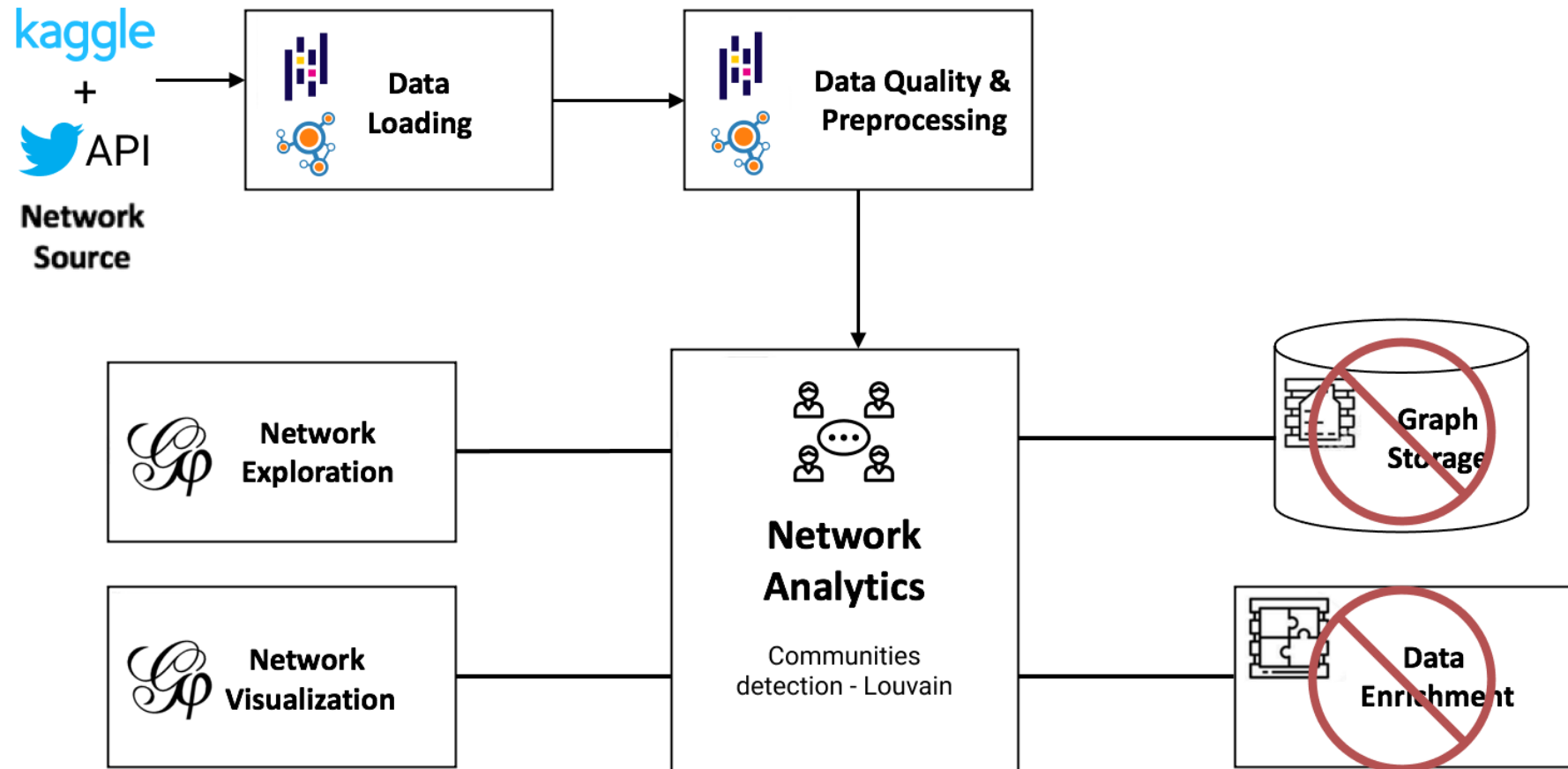# Social Media Analytics

Project Presentation
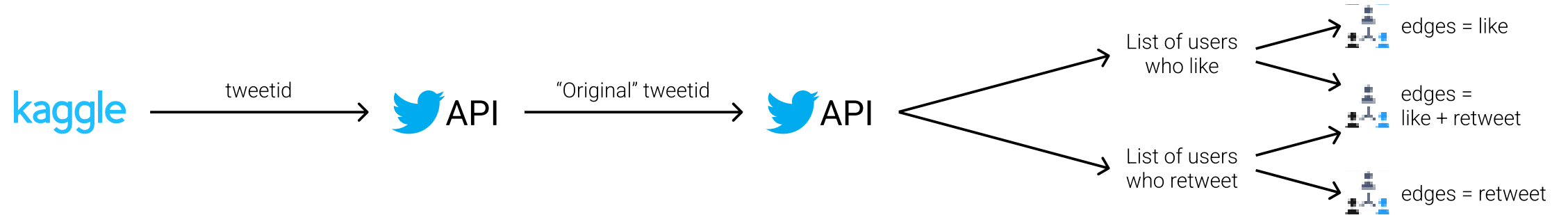
Banti Filippo
Broillet Christophe
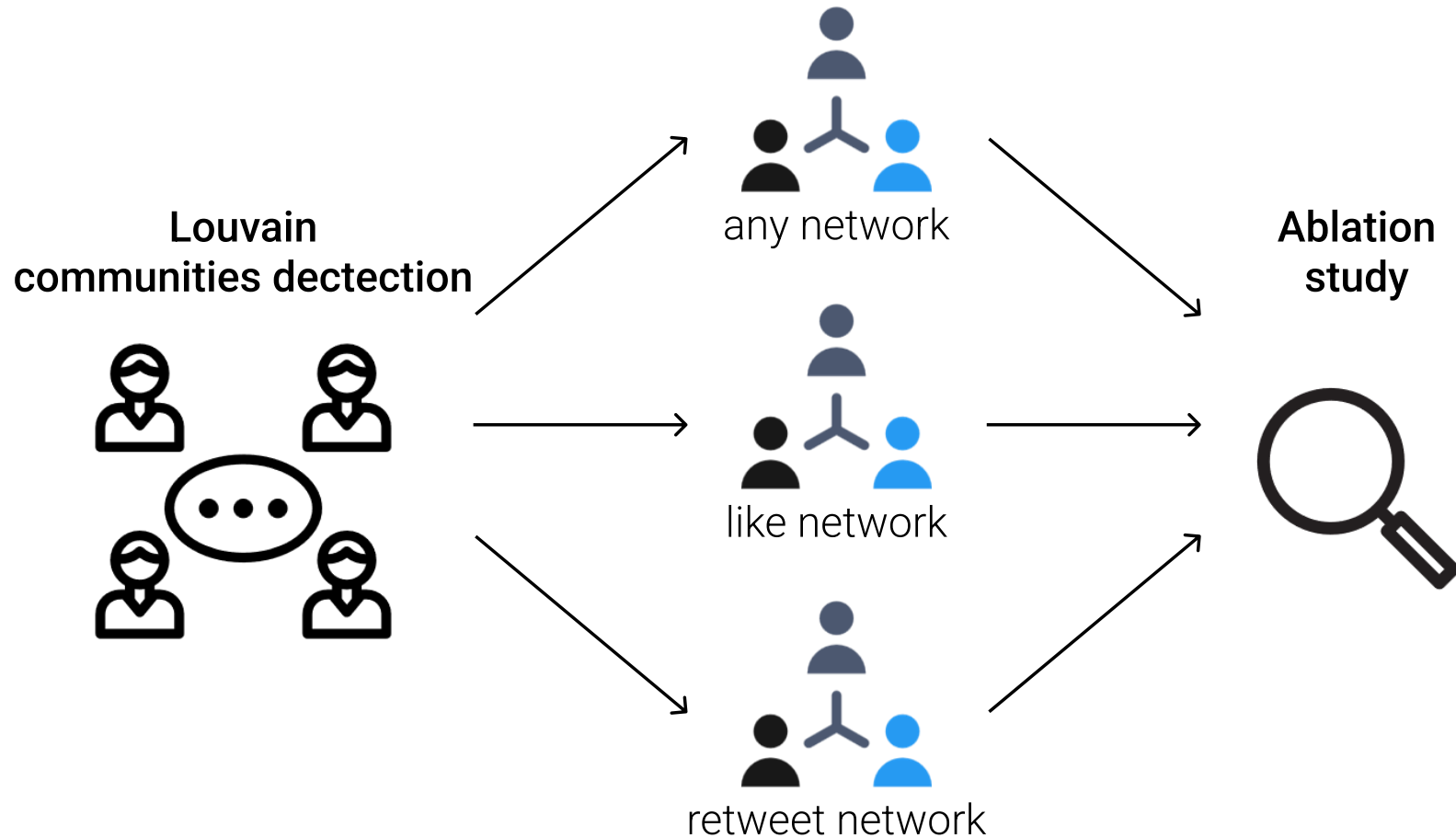Luu Quoc Bao

Overview of the project architecture

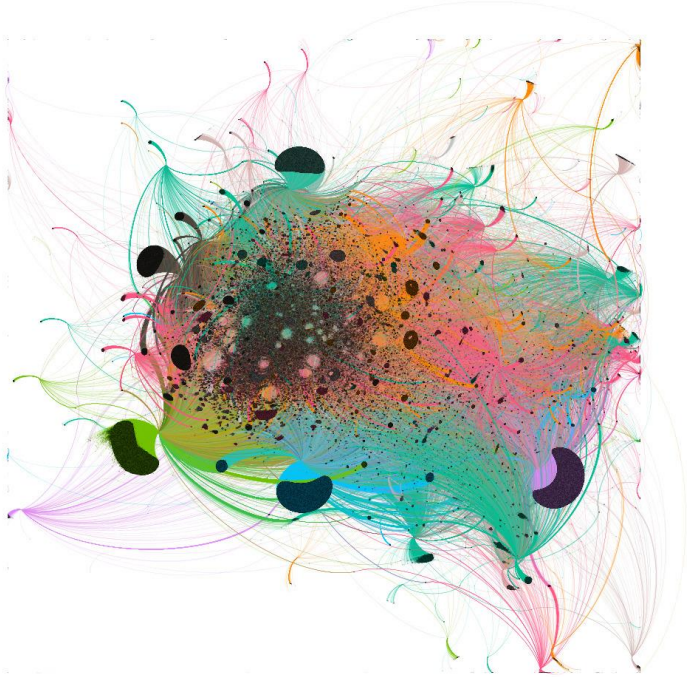# Data Loading & Preprocessing

kaggle → tweetid → 🐦 API → "Original" tweetid → 🐦 API →

List of users who like → edges = like

List of users who like → edges = like + retweet

List of users who retweet → edges = retweet

Network Analytics - overview

Louvain communities dectection

any network

like network

retweet network

Ablation study

Network Visualization

Any network

Like network
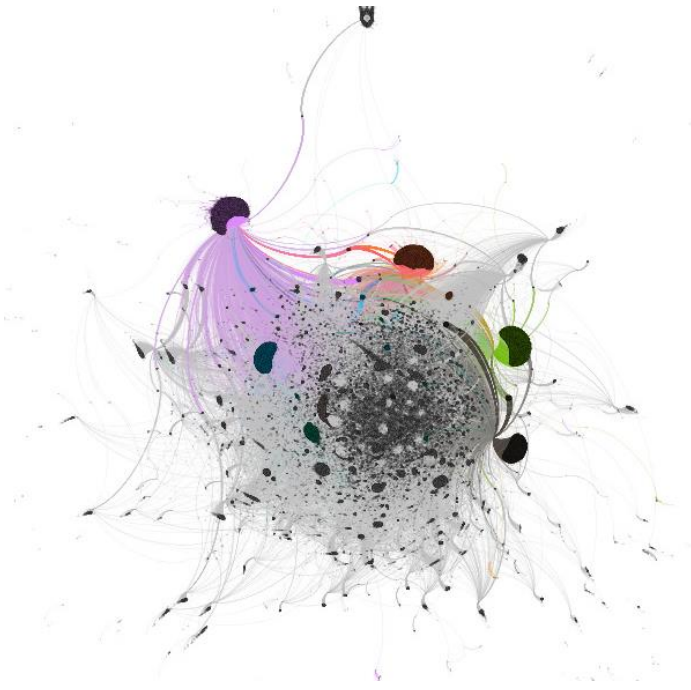
Retweet network

Thank you for your attention

Network Source

34.44M tweets

About the ongoing
war in Ukraine

No network yet

Only a list of tweets
without any links

kaggle

17 attributes

tweetid, userid,
username, acctdesc,
..., extractedts

Twitter API

To retrieve the data
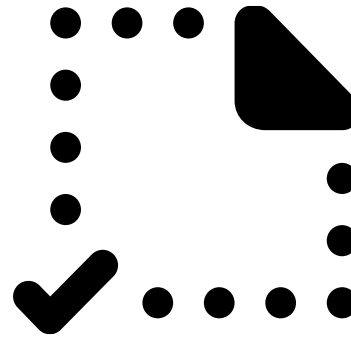necessary to form
the network
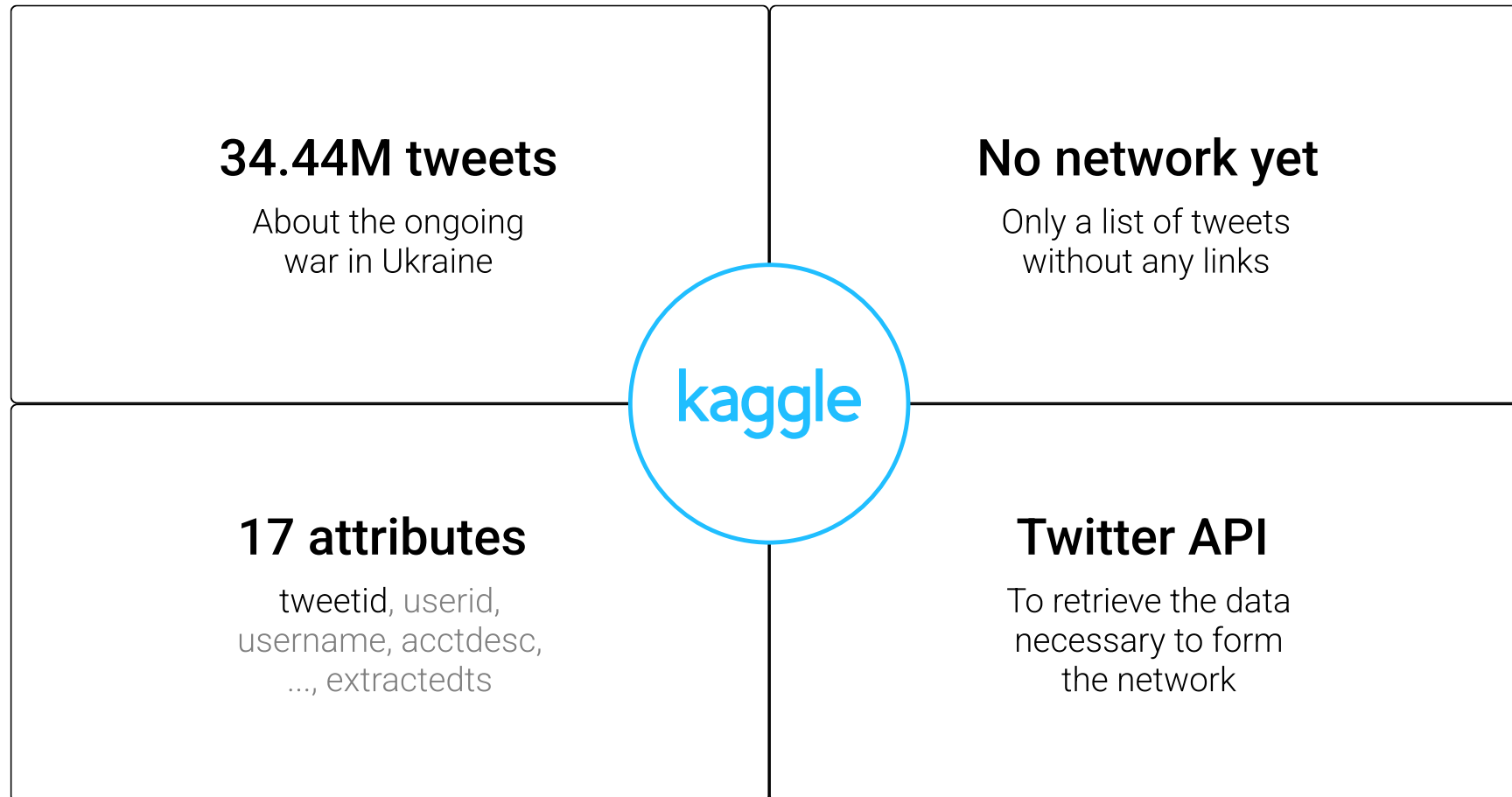
Data Quality

Deleted tweets and users

Missing data

Network Exploration

	Likes only	Retweets only	Both likes and retweets
Number of nodes	2669418	520842	2790891
Number of edges	4245301	787529	4466776
Average degree	1.59	1.51	1.60
Average clustering coefficient	0.405	0.048	0.405

Table 6: Some statistics for the three graphs
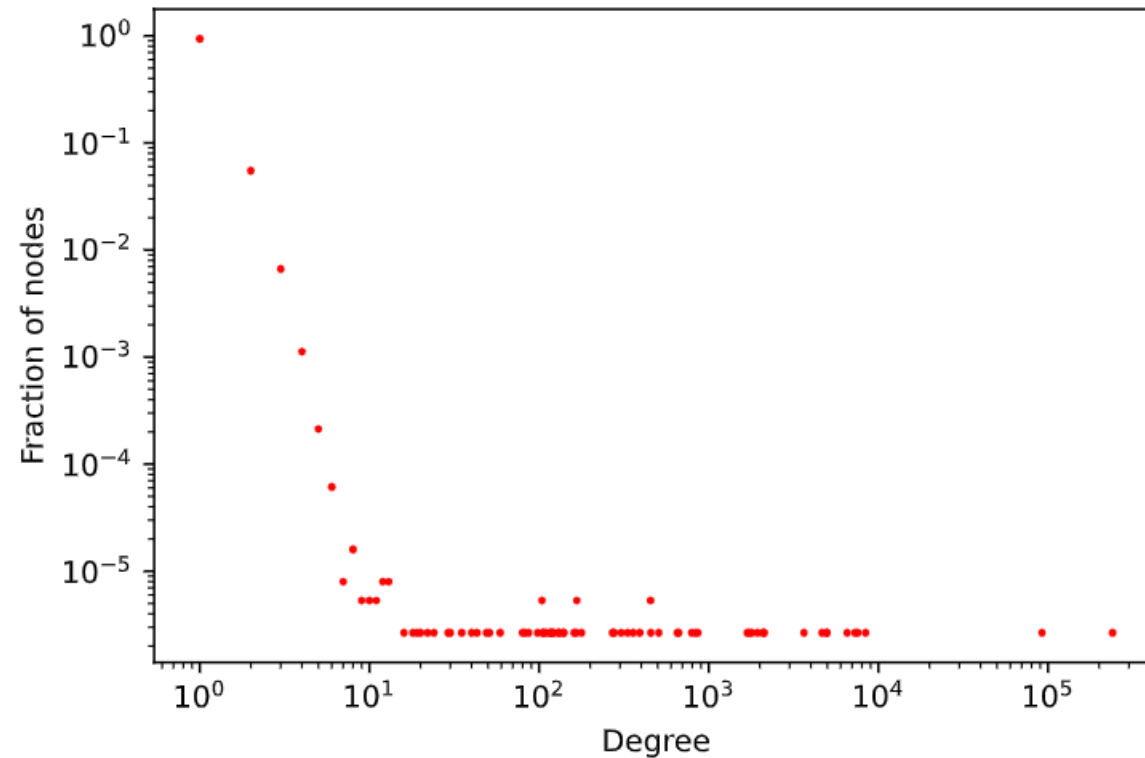
Network Exploration

Figure 2: Degree distribution for the network containing both likes and retweets.

Network Analytics - results

Passage	Likes only	Retweets only	Both likes and retweets
1	742	660	770
2	130	166	122

Table 3: Number of detected communities from Louvain, for the three different networks

Network Analytics -Comparison

	Likes	Retweets	Any
Likes	-	7.69%	27.77%
Retweets		-	8.69%
Any			-

Table 4: Intersection over union for authors in the biggest community

	Likes	Retweets	Any
Likes	-	0.00%	5.55%
Retweets		-	0.00%
Any			-

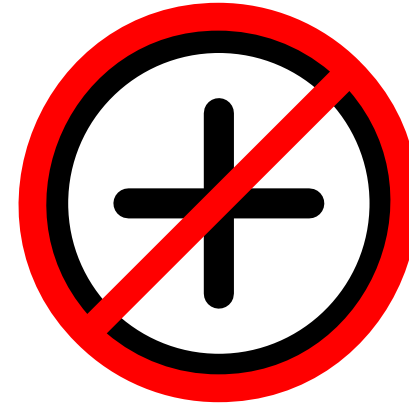Table 5: Intersection over union for authors in the second biggest community

Graph storage

- No added value
- CSV format was sufficient and easier to work with
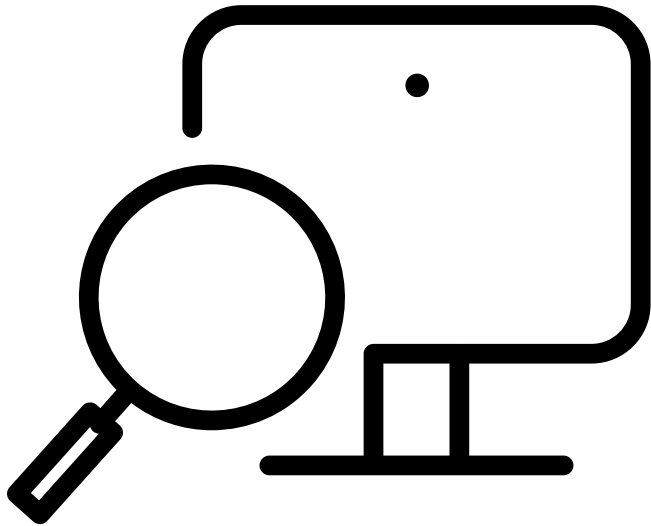- CSV format make the transfer from NetworkX/Pandas to Gephi easy
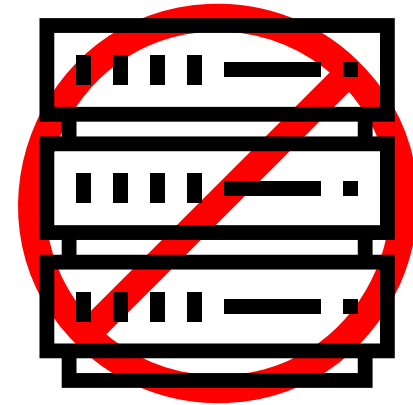
Data Enrichment

- Enough data (too much) at the source already

- Limited by the tweeter API

- No real added value

Conclusion

Results

Outlook & future work